

PREDICTING CARDIOVASCULAR DISEASE BY USING MACHINE LEARNING

Abstract

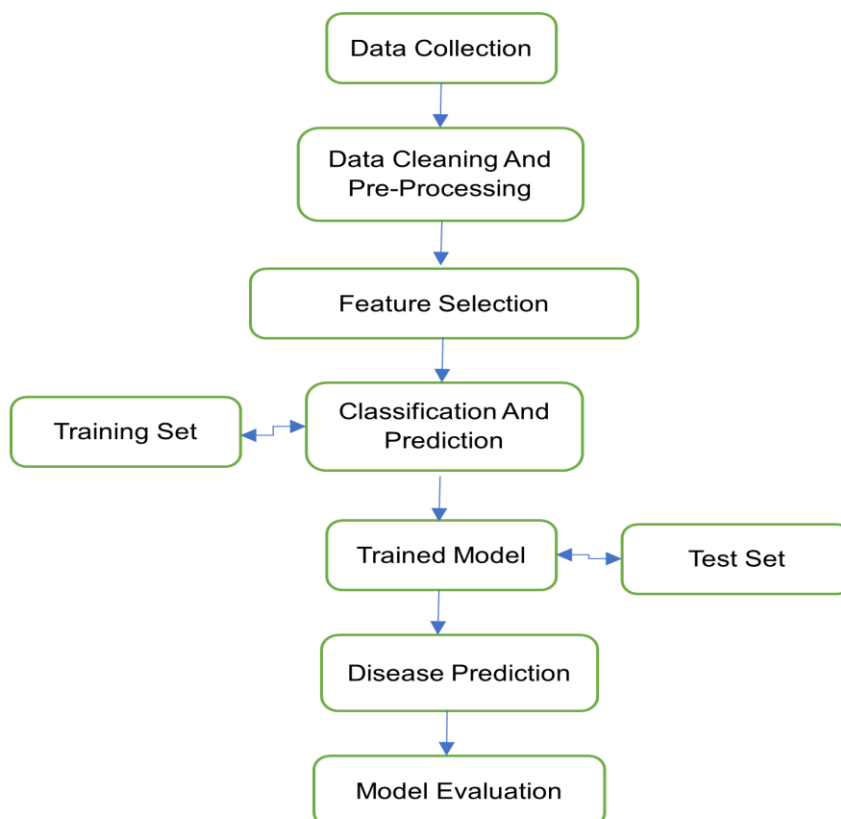
This study focuses on predicting cardiovascular disease using machine learning, particularly the KNN classifier, and leveraging the Cleveland heart disease dataset. By emphasizing the importance of feature selection, the research highlights the transformative role of machine learning in improving healthcare practices for more effective disease prediction and management.

Dataset

The Cleveland Heart Disease dataset from the UCI (University of California, Irvine) machine learning repository was utilised for this study. The dataset has 303 entries with 14 characteristics.

<https://archive.ics.uci.edu/dataset/45/heart+disease>

Methodology



Conclusion

In this study, a machine learning model was developed to predict cardiovascular disease using the Cleveland heart disease dataset. With 303 observations and 14 attributes, the model aimed to classify the presence or absence of heart disease using three algorithms: Decision Tree, k-nearest neighbour (KNN), and artificial neural networks (ANN). The dataset underwent preprocessing, addressing issues like missing values, outliers, and class imbalance. Feature selection, guided by the correlation matrix, identified the most relevant features. Two models were built for each algorithm, one with all attributes and another after feature selection. Evaluation metrics showed that after feature selection, KNN outperformed Decision Tree and ANN, demonstrating the significance of tailored feature inclusion in enhancing predictive accuracy. The study concludes that machine learning approaches effectively predict cardiovascular illness, emphasizing the importance of data quality and relevant features for optimal model performance.